Published in final edited form as:
Neuropsychologia. 2014 September ; 0: 375–389. doi:10.1016/j.neuropsychologia.2014.04.014.

A Neural Network Model of Individual Differences in Task Switching Abilities

Seth A. Herd[1], **Randall C. O'Reilly**[1], **Tom E. Hazy**[1], **Christopher H. Chatham**[1,2], **Angela M. Brant**[1,3], and **Naomi P. Friedman**[1,4,*]

[1]Department of Psychology and Neuroscience, University of Colorado Boulder, 345 UCB, Boulder, CO 80309, USA

[4]Institute for Behavioral Genetics, University of Colorado Boulder, 447 UCB, Boulder, CO 80309, USA

Abstract

We use a biologically grounded neural network model to investigate the brain mechanisms underlying individual differences specific to the selection and instantiation of representations that exert cognitive control in task switching. Existing computational models of task switching do not focus on individual differences and so cannot explain why task switching abilities are separable from other executive function (EF) abilities (such as response inhibition). We explore hypotheses regarding neural mechanisms underlying the "Shifting-Specific" and "Common EF" components of EF proposed in the Unity/Diversity model (Miyake & Friedman, 2012) and similar components in related theoretical frameworks. We do so by adapting a well-developed neural network model of working memory (Prefrontal cortex, Basal ganglia Working Memory or PBWM; Hazy, Frank, & O'Reilly, 2007) to task switching and the Stroop task, and comparing its behavior on those tasks under a variety of individual difference manipulations. Results are consistent with the hypotheses that variation specific to task switching (i.e., Shifting-Specific) may be related to uncontrolled, automatic persistence of goal representations, whereas variation general to multiple EFs (i.e., Common EF) may be related to the strength of PFC representations and their effect on processing in the remainder of the cognitive system. Moreover, increasing signal to noise ratio in PFC, theoretically tied to levels of tonic dopamine and a genetic polymorphism in the COMT gene, reduced Stroop interference but increased switch costs. This stability-flexibility tradeoff provides an explanation for why these two EF components sometimes show opposing correlations with other variables such as attention problems and self-restraint.

[*]Corresponding Author. Postal Address: Institute for Behavioral Genetics, University of Colorado Boulder, 447 UCB, Boulder, CO 80309, USA. Naomi.friedman@colorado.edu. Phone: 1-303-735-4457.

[2]Present address: Department of Cognitive, Linguistic, & Psychological Sciences, Brown University, Providence, RI 02906, USA

[3]Present address: Department of Psychology, Pennsylvania State University, State College, PA 16802, USA

Keywords

executive control; set shifting; computational model; genetics

1. Introduction

Understanding how people switch tasks (e.g., how the brain switches attention between a conversation and oncoming traffic) has obvious relevance in itself. In addition, detailed exploration of switch costs (the extra time it takes to perform a task when switching from a different task) has provided numerous insights into the mechanisms of human executive function (EF) (see reviews by Kiesel at al., 2010; Monsell, 2003; and Vandierendonck, Liefooghe, & Verbruggen, 2010). A number of computational models have been proposed as explanations of the computational and neural mechanisms of task switching. Although these models have elucidated the sources of many task switching findings, they have tended to focus on mean effects that occur across subjects, without considering patterns of individual differences that might shed light on the mechanisms involved. In this paper, we present a biologically based neural network model of task switching that can explain patterns of individual differences in terms of specific neural mechanisms.

Importantly, this model is integrated with other models of executive tasks and so can elucidate what factors distinguish vs. unify switching abilities from or with other executive abilities. We focus on the largely unanswered question of why task switching abilities are separable from other EF abilities (such as response inhibition and working memory updating/capacity) in terms of individual differences (e.g., Miyake et al., 2000). Though an individual’s performance on switch tasks correlates with other measures of EF such as performance on the antisaccade or Stroop tasks, task-switching scores also capture unique variance: Different switch tasks correlate more closely with each other than they do with other EF tasks. This unique variance not only appears to be influenced by separate genes (Friedman et al., 2008), but it also appears to show some trade offs with general executive control; those who are better at switching, controlling for other EF abilities, seem to show more behavioral problems such as more attention and externalizing problems and lower self-restraint (Miyake & Friedman, 2012). Our model suggests variations in several neural mechanisms that may explain this separation of and tradeoff between different aspects of EF.

A novel aspect of this study is that we vary many biologically based model parameters over a broad range to simulate individual differences within the normal range, and in some cases extending into ranges that would be considered pathological. This approach goes beyond typical manipulations in the computational literature that involve, for example, lesioning parts of the model to ascertain their necessity for performing a task or to simulate specific neurological insult. By focusing on more graded manipulations of key parameters, we simulate individual differences in model performance that can explain observed patterns of correlations and anti-correlations in the literature. We focus on simulating qualitative rather than quantitative patterns. Moreover, we focus on explaining results found with normal population samples, in which individuals may fall on a spectrum of ability (and disorder), with the extremes capturing individuals who might meet criteria for a diagnosis. Our model suggests variations in several neural mechanisms that may explain this separation of and tradeoff between different aspects of EF. First, however, we situate our model in relation to previous models and theories of task switching.

1.1. Previous Models of Task Switching

Theoretical and computational models of task switching have focused on two possible sources of switch costs (Kiesel et al., 2010; Vandierendonck et al., 2010). While these possibilities were considered competing explanations in early work, more recent theoretical work holds that both are likely true (Vandierendonck et al., 2010). The first possibility is that switch costs reflect time needed to resolve interference from prior conflicting task sets (rules for mapping stimuli to responses on a given task, often thought to be instantiated neurally as representations of continuously firing neurons in prefrontal cortex or PFC). This task-set interference explanation includes, in some formulations, activation triggered by previous stimulus-task associations (i.e., “task-set inertia”; Allport, Styles, & Hsieh, 1994). The second possibility is that switch costs reflect a time cost for active reconfiguration of task sets (e.g., Rogers & Monsell, 1995) arising, in some accounts, from retrieving the task set from long-term memory (Altmann & Gray, 2008, Logan & Schneider, 2010). Task sets are representations guiding performance of a particular stimulus-response mapping when stimuli have been linked to many possible responses. They likely include rules about categorization of particular stimuli, response mappings, attention orientation, and response threshold (Vandierendonck et al., 2010). Mechanistically detailed models of task switching and executive control usually hold that these task set representations consist of persistent neural firing in PFC (e.g., Cohen, Dunbar, & McClelland 1990; Herd, Banich, & O’Reilly, 2006; Reynolds, Braver, Brown, & Van der Stigchel, 2006). This mechanism is based on abundant empirical evidence from monkey electrophysiology in working memory tasks, and human neuroimaging during task switching and other EF tasks. Most abstract mathematical models are also consistent with this hypothesis (Logan & Gordon, 2001; Meiran, Kessler, & Adi-Japha, 2008; Sohn & Anderson, 2001).

Although both explanations of switch costs are likely correct in part, Vandierendonck et al. (2010) point out that an integrated view has not been implemented in a computational model (although Brown, Reynolds, & Braver, 2007, have come close with a model that incorporates a task-set representation system and interference control). Here we build on prior work by modeling task switching with a biologically based model of working memory. This model produces switch costs through both task set reconfiguration (i.e., updating the contents of working memory with new task set information based on a cue) and interference resolution (through top-down biasing and competitive inhibition). In this account, the two sources of costs are inextricably intertwined (at the level of task set representations; interference in stimulus-response mappings is separable, as we discuss later): Reconfiguring a task set takes time in part because of interference with the previous task set.

The current model is consistent with previously proposed proximal neural mechanisms of cognitive control, as explored in some detail in neural network models (Cohen, et al., 1990; Herd et al., 2006). However, we further explore their origins and interactions with other brain mechanisms crucial to making cognitive control flexible and appropriate to each situation. These earlier models show how a task set representation, in the form of persistent neural firing in prefrontal cortex (PFC), influences the neural processing that happens in other brain regions. A variety of common neural learning mechanisms (both error-driven and associative in nature) can create strong connections between neurons in the prefrontal task set and those in motor and parietal areas, thereby causing them to drive those neurons that carry out the correct mappings necessary for that task whenever the task set is active.

Crucially, those models consider task sets to be a kind of working memory. By using a well-developed model of working memory and training it to perform task switching, we follow and test this hypothesis. However, most of those previous neural network task switching models simply assumed that a task set representation is appropriately maintained during the performance of a task – that is, a specific set of prefrontal neurons remains active while a given task is performed, and that set switches when the task to be performed switches. Only a few models have included realistic mechanisms that learn task sets, (e.g., Collins & Frank, 2013; Rougier, Noelle, Braver, Cohen, & O'Reilly, 2005), and these studies do not investigate resulting switch costs in detail. Here, we explore a set of brain mechanisms that could appropriately perform that type of switching, including learning task set representations, and how to switch them appropriately. The inclusion of these mechanisms gives our model enough depth to explain not only mean effects across the population, but individual differences in task performance.

1.2. Individual Differences in Task Switching are Separable From Other Cognitive Control Abilities

Previous work has repeatedly demonstrated that EFs can be characterized as a family of related but separable abilities, both at behavioral and neural levels (e.g., Collette et al., 2005; Friedman et al., 2006; Hedden & Yoon, 2006; Lehto, Juuja rvi, Kooistra, & Pulkkinen, 2003; Miyake et al., 2000). That is, these abilities show unity and diversity (Teuber, 1972; Miyake et al., 2000). Here we explore some possible mechanistic underpinnings of two of these factors, one common to many or most EF tasks, and one specific to switching tasks. These two factors appear to be somewhat anti-correlated, so that those better at EF in general show a greater RT difference between switch and repeat trials in two-task switching paradigms than those who are weaker on general EF (although they still tend to be faster in both conditions in absolute terms). As described below, the evidence for this hypothesis is indirect but strong.

Friedman et al. (2008) reported a large twin study of three commonly examined EFs, measured with latent variables based on nine tasks: response inhibition (Inhibiting: antisaccade, Stroop, and stop-signal tasks); working memory updating (Updating: *n*-back, letter memory, and keep-track tasks); and task switching (Shifting: color-shape, number-letter, and category switch tasks). Consistent with prior literature, they found that these three factors were correlated but separable. Importantly, results of latent variable twin modeling indicated that this unity and diversity was primarily genetic in origin. Specifically, these three abilities were all influenced by a highly heritable common factor, but the Updating and Shifting components were also influenced by separate genetic factors. Friedman et al. reconfigured their model to capture the variance common to all the tasks (Common EF) and specific to updating and shifting tasks, Updating-Specific and Shifting-Specific, respectively. There was no specific variance for the Inhibiting component, suggesting that at the level of individual differences, the common factor captured all the variance in inhibiting abilities, a finding that has since been replicated in other independent datasets (see Miyake & Friedman, 2012). Their resulting Unity/Diversity model is shown in Figure 1, in which performance on shifting tasks are influenced by both the Common EF factor and the Shifting-Specific factor.

Reparsing the correlated components of Inhibiting, Updating, and Shifting into these orthogonal components of Common EF, Updating-Specific, and Shifting-Specific has resulted in the important insight that that some variables tend to show opposite correlations with the Common EF factor and Shifting-Specific factors. Table 1 provides a summary of these results for the Longitudinal Twin Sample, taken from published studies or re-analyses of published studies. The table also includes genetic correlations when available, which are the correlations between just the genetic variance in each variable (environmental correlations were generally not significant).

As this table shows, both cognitive and non-cognitive variables show a trade-off in their correlations with Common EF and Shifting-Specific abilities; however, this pattern only emerges when shifting ability is separated into the Common EF and Shifting-Specific components. For example, Friedman, Miyake, Robinson, and Hewitt (2011) reported that full-scale Wechsler Adult Intelligence Test (WAIS IQ; Wechsler, 1997) was positively correlated with Inhibiting, Updating, and Shifting abilities, but the correlation with Shifting was quite small (r = .16). However, when the same data were analyzed with the Unity/Diversity model, IQ positively correlated with Common EF and Updating-Specific, but significantly negatively correlated with Shifting-Specific. These results indicate the low correlation with Shifting was due to the fact that IQ is positively related to the Common EF part of Shifting, but negatively related to Shifting-Specific. Hence, the total correlation with Shifting, which is a combination of these two factors, was low. This same trade-off pattern can be seen in several non-cognitive phenotypes related to externalizing disorders and behavior problems, measured at multiple earlier times in the longitudinal sample, shown in Table 1. These measures include teacher-rated attention problems (Friedman et al., 2007), Behavioral Disinhibition at two ages (a latent variable that captures externalizing behaviors including ADHD, substance use, conduct disorder, and the personality dimension of novelty seeking; Young et al, 2009), and even very early self-restraint, measured at ages 14 to 26 months (Friedman et al., 2011).[1]

Taken together, these data suggest that there are opposing mechanisms that tend to make some variables, particularly behavioral problems, show negative correlations with Common EF but positive correlations with Shifting-Specific. Although these two components are by

[1] It is important to note that we do not always find this strong pattern of trade-offs. For example, Friedman et al. (2009) found that the slope of growth in sleep problems from ages 4 to 16 years was negatively correlated with Common EF (r = −.29), but was not significantly correlated with Shifting-Specific (r = .12). Similarly, in unpublished data, we have found the number of past week depression symptoms, as measured by the Center for Epidemiological Scales – Depression (CES-D; Radloff, 1977), showed a significant negative correlation (r = −.28 phenotypic and r = −.38 genetic) with the Common EF factor but did not significantly correlate with Shifting-Specific (r = .07 phenotypic and r = .07 genetic).

necessity uncorrelated in the Unity/Diversity model (because Shifting-Specific is essentially a residual of Common EF), their trade-off becomes apparent when a third variable is added to the model. These patterns have led us to postulate some mechanisms that might underlie these factors and show such a trade-off. In this paper, we show that manipulations of these mechanisms within the models do indeed result in opposite patterns of results for an inhibition task (a four-response, manual variant of the common Stroop (1935) color-naming vs. word-reading task) vs. a switching task.

Specifically, we propose that individual differences in the Common EF factor in large part reflect the ability to actively maintain goals and goal-related information, often in the face of interference (Friedman et al., 2008), and to use these goals to bias ongoing processing (i.e., top-down attention and task control). This basic ability, which has been frequently posited as a key element of executive control and frontal lobe functioning (e.g., Engle, Kane, & Tuholski, 1999; Miller & Cohen, 2001), is necessary for all types of EF tasks. It may also be particularly influential in response inhibition tasks and other tasks that involve conflict (which many of our EF tasks do), a view consistent with views of inhibitory control as a by-product of goal maintenance (Chatham et al., 2012; Herd et al 2006; Kane & Engle, 2003; Morton & Munakata, 2002; Munakata et al., 2011). This strong influence of goal maintenance on response inhibition tasks may explain why Friedman et al. (2008) found no inhibiting-specific variance in addition to Common EF explaining the correlations between the response inhibition tasks. Within the computational model, we show how individual differences in this ability are affected by two example manipulations that affect the influence of maintained PFC task representations on posterior processing: the gain parameter of the units in the PFC layer, and the strength of the connection between the PFC layer and posterior hidden layers.

In contrast, we hypothesize that the Shifting-Specific component reflects flexibility—ease of transitioning to new task set representations. To put it differently, poor performance on the Shifting-Specific component may reflect “stickiness” (Altamirano, Miyake, & Whitmer, 2010; Blackwell, Chatham, Wiseheart & Munakata, this issue), a tendency for goal representations to stick around after they are no longer relevant (i.e., in the absence of their continued *active* maintenance). In our model, this ability primarily involves the replacement of no-longer-active goals. We have explored two mechanisms that influence individual differences in this trait: recurrent connection strength in PFC, and the degree of clearing of old representations when gating occurs (clearing on gating).

1.3. PBWM Framework and Model Predictions

Our starting point for the modeling work was a well-developed neural network model of working memory: PBWM (Prefrontal cortex, Basal ganglia Working Memory (Frank, Loughry, & O’Reilly, 2001; Frank, Seeberger, & O’Reilly, 2004; Hazy, Frank, & O’Reilly, 2006, 2007; O’Reilly, 2006; O’Reilly & Frank, 2006). In part, our choice of a working memory model served as a test of the theory that cognitive control is simply one function of working memory mechanisms (Deco & Rolls, 2005; Engle et al., 1999; O’Reilly, Braver, & Cohen, 1999). In the view we present here, cognitive control and working memory use isomorphic brain mechanisms: Cognitive control arises from working memory representations of goals. These goal representations exert useful biasing effects that give rise to attention (Desimone & Duncan, 1995) and flexible stimulus-response mappings (Cohen et al., 1990, Herd et al., 2006).

The PBWM series of models is based on a wealth of physiological data, and incorporates a good deal of physiological detail. We used this detailed model of working memory to create specific hypotheses about the brain mechanisms underlying different aspects of EF, and to explore the degree to which a model intended to explain working memory could explain EF more generally. We propose links from these physiological hypotheses to the level of psychological constructs (Figure 2). Specifically, we examined the different mechanisms called upon to rapidly switch between tasks versus those needed to maintain strong representations that support all EF tasks. We discuss these distinctions primarily in the context of the Unity/Diversity model of EF (Friedman et al., 2008; Miyake & Friedman, 2012); however, this idea of a flexibility/stability tradeoff in goal-directed behavior has been discussed by a number of researchers (e.g., Bilder, 2012; Cools, 2012; Goschke, 2000; Tunbridge, Harrison, & Weinberger, 2006).

The Common EF versus Shifting-Specific constructs map relatively well to particular components of the PBWM model (although that mapping is not perfectly one-to-one, as addressed in the discussion). PBWM focuses on the need to selectively and appropriately “gate” information into working memory. Persistent neural firing in PFC over delay periods has been shown to reflect relevant items, whereas distracting items cause brief but not persistent activity (Miller, 2000). Some mechanism must appropriately select which information is maintained. The PBWM framework proposes that such selection is performed by a complex interplay between the PFC and the basal ganglia (BG), directly analogous to the well-studied interaction between sensorimotor cortex and BG in selecting motor actions.

While the BG provides the “store or don’t” signal, the PFC is responsible for actually maintaining the selected information. The BG input dramatically strengthens maintenance, but the PFC by itself also has an intrinsic ability (and therefore tendency) to maintain any and all representations it forms for a short period. While this automatic maintenance or “stickiness” of representations in PFC is helpful in many cases, these intrinsic cortical mechanisms also slow down the switch from one representation to the next. In the case of task switching, this slowed transition can slow task performance. Thus, the several neural factors that contribute to a tendency to maintain information in PFC (regardless of the gating decision from BG) should contribute to worse switching performance. Neural factors that are controlled by BG gating should not, on the other hand, slow down switching, as such gating happens rapidly. Such factors should contribute to the Common EF component. One exception is factors that are triggered by gating, but serve to actively clear out old representations; these should also contribute to the Shifting-Specific component (see Figure 2).

2. Methods

2.1. General Model Architecture

Our model, shown in Figure 3, was constructed based on the Leabra neural network modeling framework (O'Reilly, 1996; O'Reilly, Hazy & Herd, in press; O'Reilly & Munakata, 2000). This framework fulfills some of the criteria laid out by Newell (1990) for a cumulative model that explains many phenomena, and so constitutes a more constrained theory. The Leabra framework has been used to model many different psychological and neurophysiological findings with the same basic set of assumptions and most of the same parameter values (see O'Reilly et al., in press, for a recent review). This model also gained additional constraints in using parameters and network architecture to model not only one task, as in most neural network work, but all nine of the tasks used by Friedman et al (2008) in a closely related modeling project (Friedman et al., 2014; see also Chatham et al., 2011).[2]

The Leabra algorithm is described in greater detail in print (O'Reilly & Munakata, 2000) and actively updated online textbooks (O'Reilly et al., 2012), and its purpose and tenets as a cumulative theory are elaborated in O'Reilly et al. (in press); its more relevant general characteristics are described here. All of these attributes are common choices in artificial neural network models, except as noted. Point neurons are used, with no explicit modeling of their spatial extent. Input from other units is summed and translated to an average firing rate (a "rate coded" approach). The effect of a unit on each of its afferents is determined by the product of its firing rate and the synaptic strength connecting the two. All such weights in the current model were "learned" in the course of training, by adapting their strength through a combination of error-driven and associative learning, a learning rule unique to Leabra. The error-driven component is based on phasic differences between an early "guess" or "minus" phase, and a late "correct answer" or "plus" phase; information from the correct answer affects the network through feedback synaptic top-down projections (which are identical in function to the excitatory feedforward projections). The model performs learning with a similar overall effect to backpropagation learning, but calculates weight changes by comparing differences in local synaptic activity, rather than the non-local and therefore biologically implausible operation employed in backpropagation learning. The error-driven component of the learning essentially drives the network to create states from inputs that are similar to those that occur in the presence of the appropriate output. The source of the correct answers needed for this type of learning can, in many cases, be conceptualized as arriving from later sensory input or from other brain areas; an actual external teacher is thus beneficial but not necessary for this type of learning. In the current training, we assume that reliable feedback on desired answers is available in some form.

The PBWM model in particular is based on extensive lesion and electrophysiological evidence from those brain areas involved in reward-motivated instrumental and working memory tasks in animals, and it is constrained by human behavioral and neuroimaging

[2] That project focuses on using the same model to capture individual differences in all nine of the tasks used by Friedman et al. (2008). It does not include discussions of benchmark findings but rather focuses on understanding the mechanisms that may lead multiple tasks within a construct to intercorrelate. Although there is some discussion of the parameter manipulations used here, it is in the broader context of understanding patterns seen across all nine tasks, including updating tasks.

studies on more complex working memory tasks in humans. The extensive work on the role of the BG and cortex in motor learning is leveraged to explain the brain mechanisms underlying working memory, because the anatomical relationships between these systems are highly similar in the areas known to be responsible for motor learning and working memory. In essence, the PFC creates representations of candidate actions, task sets, or memory items, and the BG decides which of these representations are “gated” to be performed or remembered. The PFC learns these representations through error-driven mechanisms, while the BG learns the final gating decisions based on predictions of reward. Those reward predictions are made by an array of subcortical systems working in concert (Hazy, Frank, & O’Reilly, 2010; O’Reilly, Frank, Hazy, & Watz, 2007). These systems produce dopamine release when new predictors of reward appear in the environment, and dips when expected rewards do not occur. These phasic changes in dopamine concentration in turn drive learning in the striatum (labeled Matrix in model illustrations for the relevant subset of striatal neurons) to perform actions and maintain memory items that predict reward. The details and further explanation of the computations of the PBWM model can be found in the relevant publications (Frank, et al., 2001, 2004; Hazy, et al., 2006, 2007; O’Reilly 2006; O’Reilly & Frank, 2006).

The model is highly similar but not identical to previously published versions, because the framework is still being actively developed. This model was developed from an off-the-shelf current version of the model. Parameters were changed only to put the model in a more sensitive parameter regime to allow investigation of individual differences, since the development version was optimized for the fastest possible learning on a small set of benchmark tasks. In setting the default levels of various parameters, we selected intermediate values so that performance would not be optimized, but could increase or decrease, thus improving our chances for obtaining individual differences with manipulations. In particular, we used default values for the parameters we manipulated that would produce performance comparable to average human performance, and not at ceiling or floor (e.g., gain parameters for the neural response functions in PFC and hidden layers were reduced from 600 to 100; the clear-on-gating parameter was reduced from 1 to 0.5).

The other important difference between this and other PBWM models is the inclusion of two hidden layers, and the persistence of neural activity in those layers between trials. These layers produce an attractor state that gives the model an increased tendency to make the same sensory-motor mappings that it did on the previous trials. This change captures the persistence of representations in premotor cortex, higher sensory areas, and parietal cortex that perform stimulus-response mappings (e.g., Curtis & Lee, 2011).

2.2. Task Details

The switch and Stroop tasks are implemented in the same model, so their architectures and parameters are identical. We determined a training schedule for each task that would lead to performance approximating average accuracy obtained in a large normal population sample (Friedman et al., 2008; 2011) at default parameter settings. The same training schedules were used for testing the benchmark findings (all tested with default parameter values) and for testing individual difference manipulations (see Section 3.2.1).

2.2.1. Switch Paradigm—The results we report here were primarily based on a color-naming/shape-naming task switching paradigm in which two classes of stimuli and manual responses are used for each task (Miyake, Emerson, Padilla, & Ahn, 2004). In the first model settling trial (*cue naming*), the task cue was presented, and the model was trained to echo the task name in a verbal output layer ("color" or "shape"). In the immediately following *task performance* trial, the task cue was not presented so that the network would rely on the working memory capacity of PFC to maintain that information. Because activity in the hidden layers (roughly corresponding to parietal cortex and other areas within the stimulus-response arc) was decayed by reducing activations by half between trials, adequate information could be recovered from the representations in the hidden layers to perform the task. However, the model consistently learned to use the gating mechanisms of PFC to maintain this information as well, and the significant effects of the manipulations of parameters in PFC provide reassurance that it plays an important role in task switching in the model, as it is known to do in people. The model was trained to produce the response appropriate to the cued task on the manual output layer (assuming a manual response on a button-box), and to produce the name of the cued task once again on the verbal layer (to scaffold the creation of a representation of the previously cued task). Importantly, the manual output responses are shared across tasks (e.g., the same output unit is used to indicate "red" and "circle," depending on the task, making these manual responses bivalent). Each behavioral trial consisted of two model trials.

The use of pre-cuing vs. presenting both task cue and stimuli at once was crucial; unlike humans, neural network models tend to learn such tasks as a flat, conjunctive mapping of task cue plus stimulus to output when both cue and stimulus are presented simultaneously. This type of learning produces fundamentally different kinds of switch costs, since tasks are not explicitly represented. Including a small working memory requirement produces a better model of human behavior. Training the network to name the task as well as to respond on the performance trial did not appear to be critical, but did allow the model to learn faster and more reliably. This procedure matches human behavior under the hypothesis that cue processing and task set preparation are separate operations from actual task performance.

In this way, our model is unlike typical behavioral experiments, in which only a single response time (RT) is available per trial; in the model, time preparing to perform the task is tracked separately from time actually performing the task. Arrington, Logan, and Schneider (2007) used a very similar design to separate cue and task switch costs in a behavioral experiment, operating under the same assumption that cue and target processing are largely serial (cf., Logan & Bundesen, 2003). They found switch effects in both the cue and target phases, and reported that responding to the cue did not disrupt normal processing. Hence, the version of the task used for this model would seem to generalize to more typical versions in which only one response is required.

The model was trained using shaping — a technique widely employed in the training of primates (e.g., Skinner, 1938; see also Krueger & Dayan, 2009, for application to computational models). It was first trained for twenty epochs on switching at random: Each task appeared equally often, with a 50% chance of switching tasks on each trial. This frequent switching training was needed to establish basic performance. Next the model was trained for 100 epochs, still performing each task equally often, but with only a 20% chance of switching tasks on each trial, and without the provision of a task cue on repeat trials. The cueing phase was eliminated on repeat trials, although the model still had to name the appropriate task for every trial, using memory traces, and of course needed to maintain that information to provide the correct manual response on conflict trials. The model was cued to the new task with every switch. This style of training emulated the more standard experience people have with task switching: Task sets tend to remain relevant for some time until a signal in the environment indicates that a different task set is needed. This training built an expectation for repeat trials. (While this style of training was not necessary for producing switch costs, it did make the costs larger, and may be more ecologically valid.) Finally, the model was trained for three epochs in the original 50% switch condition, but with the task cue presented during the response portion of the trial as well as the cue period (emulating practice trials at the laboratory task). The model was then tested for ten epochs without learning (no change in any synaptic weights). All reported results are from those test epochs.

RTs were modeled as the number of cycles it took for an output unit to reach a threshold of 0.5 activation. Only RTs for trials that ended with correct responses were analyzed (additionally, we eliminated trials following errors, as the correct task set may not have been achieved on the prior trial. This elimination of trials following errors is common in analyzing switch task data; e.g., Friedman et al., 2008).

2.2.2. Stroop Paradigm—We also modeled the Stroop task as an example of the response inhibition tasks found to load only on the Common EF component (Miyake & Friedman, 2012). The Stroop paradigm was nearly identical to that for the switch task (with two tasks: word reading and color naming), but training involved a higher probability of the word-reading task vs. the color-naming task to build a prepotent response to read the words. Four responses and stimulus variations per dimension were used. The Shape layer was used for word-reading inputs and the Color layer for color-naming inputs; words were locally coded (a single unit for each) in the input layer in the same manner as color and shape, to provide a simple mock-up of the input from sensory processing areas. For the Stroop task, each model was trained for 20 epochs of basic training, with equal numbers of word-reading and color-naming trials. In this phase, all trials were “neutral” (inputs from the irrelevant modality were neither conflicting nor congruent, as they had no trained response mapping). Next, the models were trained for another 40 epochs with asymmetric training, such that 33% of trials were color naming. In this phase, there were 40% neutral trials, 40% congruent trials, and 20% incongruent trials. The model was then trained for 87 epochs (a number chosen arbitrarily) on 80% word-reading trials, with no cue given for repeated trials. This training forced the model to encode word reading in the hidden layers, with PFC task sets engaged primarily for the rare color naming trials. This phase also used 40% neutral, 40% congruent, and 20% incongruent trials for both tasks. This training was followed by three epochs of training on color naming alone (simulating practice on the laboratory color-naming Stroop task) with equal numbers of neutral, congruent, and incongruent trials. The model was then tested on color naming in this same condition, without further learning.

3. Results

3. 1. General Task Switching Results (Benchmark Findings)

In this section we show that our model reproduces several key benchmark findings (Vandierendonck et al., 2010), based on 100 runs. Although the model did not permit investigation of every task switching effect in the literature (e.g., the "backward inhibition" effect observed in tasks that require switching between three subtasks; Mayr & Keele, 2000; or cue switch vs. task switch costs observed when multiple task cues are used), it does permit investigation of most findings. Specifically, we examined whether the model showed switch costs, interference costs, mixing costs, and, because our model can perform both Stroop and switch tasks, asymmetric switch costs. These results were obtained using default values of all parameters, in contrast to the individual difference manipulations discussed in Section 3.2.

3.1.1. Switch Costs—First and most importantly, our model shows a switch cost. Accuracy was near ceiling, so there was no significant switch cost in accuracy for the cue naming trials (switch 98.6% vs. repeat 99.5%, *t*(99) = −1.44, *p* = .152) or task performance trials (switch 98.7% vs. repeat 99.6%, *t*(99) = −1.33, *p* = .186), although the effects were in the right direction.[3] However, there were robust costs in the RTs. The total switch cost (sum of the cue naming and task performance RTs) was 2.28 cycles (SE = .065; *t*(99) = 34.99, *p* < .001). Cost was significant for both the time to name to the cue in switch vs. repeat trials (M = 2.11 cycles, SE = .056; *t*(99) = 37.41, *p* < .001), and the time to categorize the stimulus after responding to the cue, (M = 0.17 cycles, SE = .040; *t*(99) = 4.21, *p* < .001). The split of the trials into cue naming and task performance can be considered analogous to administering trials with long cue-to-stimulus intervals that allow the subject to prepare: In this situation, the switch cost for the task performance trial can be considered analogous to the "residual" switch cost that is observed even when subjects have ample time to process the task cue. The fact that the switch cost for task performance trials was smaller than the total cost is consistent with the typical finding of a preparation benefit; the fact that the task performance trials still showed a significant switch cost despite preparation is consistent with the phenomenon of residual switch cost (Kiesel et al., 2010, Monsell, 2003; Vandierendonck et al., 2010).

Our residual switch cost is quite small in proportion to total RTs for the model (<10% of the total switch cost). Typically, residual switch costs are larger (e.g., in our own research we have found residual switch costs approximately 25% to 40% the magnitude of regular switch costs; Friedman & Miyake, 2004). The small size of the residual switch cost in our model might be interpreted as providing some support for each of the two competing theories about residual switch costs. On the one hand, the small residual switch costs might be considered consistent with theories in which residual switch costs largely reflect a "failure to engage" on a small proportion of the trials (De Jong, 2000). The scope of our model does not extend to simulating such occasional failures to prepare (c.f. Reynolds, et al., 2006); the model

[3] Switch costs in accuracy are often present in behavioral data. In our own data (unpublished), we have found that small but significant mean switch costs in accuracy are not reliable in terms of individual differences, most likely because of restricted range due to near-ceiling performance. Because most of the results that we aim to model are based on RT data, we focus on those effects.

cannot fail to prepare by focusing on task-irrelevant stimuli or thoughts, as humans do. Thus, we are capturing less than half the full magnitude of typical residual costs, possibly because our model never fails to engage. On the other hand, the fact that we still find some significant residual switch cost (though very small), despite no failures to engage, suggests that interference effects may also contribute to this residual cost. Thus, it seems plausible that in human performance, residual switch costs could be due to both interference and failure to engage, consistent with the proposed theoretical integration of Vandierendonck et al. (2010).

3.1.2. Mixing Costs—In many switch tasks, although repeat trials are faster than switch trials in mixed blocks (those that include both tasks), such repeat trials are still slower than trials in pure-task blocks. This “mixing cost” has been attributed to a number of factors, including increased interference of tasks in mixed lists and the requirement to maintain multiple task sets and response mappings (Kiesel et al., 2010). Rubin and Meiran (2005) found evidence supporting the hypothesis that mixing costs reflect resolution of task ambiguity induced by bottom-up activation of competing task sets. This ambiguity may be influenced by the subject’s knowledge about the presence of mixing vs. not, which is usually given in the task procedure. That is, in single-task blocks, the subject knows the other task will not be relevant, so it may be easier to maintain a stronger task set of the relevant task than when subject knows the task will be constantly be switching. The opportunity to practice a consistent set of stimulus-response mappings is likely also a contributing factor for faster performance when blocks of trials are not mixed.

If practice is a contributing factor, we might expect to see large mixing costs in this particular paradigm only when the model has experienced learning with probabilities (so the model could adjust its weights for task probability). Thus, we examined whether the model would show mixing costs in two different ways to shed light on the mechanisms in the model. First, we trained the model with three epochs of learning on switch trials (as in the standard learning paradigm), then we ran a non-learning test on both switch blocks and single-task blocks. Thus, the model’s weights should be set as if it were always expecting switch trials. Nevertheless, we did see a small mixing cost when comparing repeat trials in mixed blocks (RT = 35.61 cycles, SE = 0.149) to trials in single-task blocks (averaged across color and shape blocks, RT = 35.38 cycles, SE = 0.151), $t(99) = 5.49$, $p < .001$. However, this slightly longer RT (0.23 cycles, SE = 0.041) on repeat trials in mixed blocks was somewhat small compared to typical mix costs (which can be about half the size of switch costs; e.g., Kray & Lindenberger, 2000), and likely reflects increased residual activation from the other task set that persisted across multiple trials (i.e., from task sets that were active two or more trials before).

To examine whether learning task probabilities play a role in the mixing costs, we used the same procedure to test single task blocks: 3 epochs of learning with the task, then 10 epochs of non-learning test. That is, the model received 3 epochs of learning with the switch blocks, then a no-learning test; 3 epochs of learning with the color task before a no-learning test; and 3 epochs of learning with the shape task before the no-learning test. In this case, we saw a much larger mixing cost when comparing repeat trials in mixed blocks (RT = 35.61 cycles, SE = 0.149) to trials in single-task blocks (RT = 33.91 cycles, SE = 0.109), $t(99) = 20.79$, p

< .001. This longer RT (1.70 cycles, SE = 0.082) on repeat trials in mixed blocks was more comparable to typical mix costs. This mixing cost was also comparable to that obtained when we allowed learning during test trials: Repeat trials in mixed blocks (RT = 34.99 cycles, SE = 0.127) were slower than trials in single-task blocks (RT = 33.66 cycles, SE = 0.108; mixing cost = 1.33, SE = 0.058), *t*(99) = 20.79, *p* < .001. Thus, it appears that the mixing cost in this model is due to some small extent on residual activation from competing task sets, but to a larger extent to consistent learning about stimulus-response mappings.

3.1.3. Response Interference Effects—A robust finding in switch tasks is that responses to stimuli that afford different responses for each task are slower than those that afford the same response — an incongruency cost. Moreover, this effect interacts with switching, such that the incongruency cost is greater for switch than repeat trials (Kiesel et al., 2010; Vandierendonck et al., 2010). We reproduce this pattern of results with the RTs for the task performance trials. Specifically, incongruent trials (RT = 21.18 cycles, SE = .139) were slower than congruent trials (RT = 16.70 cycles, SE = 0.092), *F*(1, 99) = 1494.80, *p* < .001. This effect interacted with switch condition, F(1, 99) = 22.64, *p* < .001, such that the incongruency cost was larger for switch trials (M = 4.68 cycles, SE = .125) than for repeat trials (M = 4.31 cycles, SE = .118). The fact that we find this effect even with a prior cue naming (analogous to a long preparation time) is consistent with the finding that task-rule congruency is not affected by preparation time (Monsell, 2003, but see also Kiesel et al., 2010).

Another effect that occurs in switch tasks is a response repetition effect in which there is a benefit for response repetitions on repeat trials, but there is no response repetition benefit, or in some studies even a response repetition cost, on switch trials (e.g., Rogers & Monsell, 1995). Our model does not reproduce this finding. Specifically, though the response repetition benefit on switch trials (M = .02 cycles, SE = .037) was not significant, *t*(99) = 0.53, *p* = .595, it was only marginally significant on repeat trials (M = .06 cycles, SE = .030), *t*(99) = 1.88, *p* = .063. The interaction between switch condition and response repetition was not significant, *F*(1, 99) = 0.48, *p* = .489. Hence, although the general pattern was present, the effect was not statistically significant. It is possible that the presence of the cue naming trials (similar to long preparation times) may have contributed to this absence of an effect. Because the task implementation included cue naming trials between each pair of task performance trials, the carry-over activation from one categorization response trial to the next decayed more than would be the case had the categorization response trials immediately followed one another. Extending the model to include adaptive feedback mechanisms in response to conflict or errors (see discussion) might also produce this effect. Finally, adding the known stronger short-term effects of learning (the well-established Short Term Potentiation (STP) phenomenon), should also increase this effect by increasing the degree to which a particular response activates the task set with which it last occurred.

3.1.4. Asymmetric Switch Costs—When individuals have to shift between tasks that have different strengths or prepotency, they show asymmetric switch costs. For example, Allport et al. (1994) found that in a Stroop task involving switching between word reading and color naming, switch costs were, somewhat counter intuitively, larger when switching to the dominant word reading task. To reproduce this finding, we tested a model that had been trained on the Stroop task (to produce a task asymmetry) with 10 epochs of cued switch trials (50% word-reading, and 50% color-naming; as with other models, it was a test with no learning that was preceded by three epochs of training with the exact task before learning was turned off). Consistent with prior results, we found that color-naming task performance trials (M = 15.05 cycles, SE = .058) were slower than word-naming task performance trials (M = 14.63 cycles, SE = .049), F(1, 99) = 49.58, p < .001, and combining cue naming and task performance trials, switch trials (M = 30.34 cycles, SE = .105) were slower than repeat trials (M = 28.52 cycles, SE = .098), F(1, 99) = 938.56, p < .001; However, there was a significant task by switch condition interaction, F(1, 99) = 17.39, p < .001, such that switch costs were larger when switching to word-reading (M = 2.06 cycles, SE = .094) than to color-naming (M = 1.58 cycles SE = .069). Consistent with the arguments of Allport et al. (1994), this effect seems to be caused by the necessity of a stronger task set for the less prepotent task; that task set is less easily replaced by the weaker task set needed to support the more commonly practiced task (the task set representation for the more-practiced task is weaker because those well-practiced stimulus-response mappings require less support from the PFC, and the error-driven learning that creates those task sets makes them as strong as they need to be to produce correct answers). The model’s relationship to other proposed explanations for the effect are considered in Section 4.2.

3.2. Individual Difference Manipulations

The reproduction of the most commonly reported effects of task switching — including task-switch and residual switch costs, mixing costs, congruency effects, and asymmetric switch costs — suggests that our model is rich and accurate enough to perform its main function: addressing the mechanisms underlying the different components of EF abilities. In this section we report the results of our parameter manipulations designed to simulate individual differences in the switch task, and also in the comparison Stroop task. We focus on comparing the effects of manipulating parameters that correspond to individual differences in EFs, which appear to be largely genetic (Friedman et al., 2008). To explore the mechanisms underlying the Common EF and Shifting-Specific components, we tested two candidate parameters for each. We demonstrate dissociations between effects on the model’s performance in the Stroop task (one of the response inhibition tasks most strongly dependent on the Common EF component in Miyake & Friedman, 2012), and effects on the switch task we focus on here. First, though, we briefly outline the procedure we used to test these parameter manipulations.

3.2.1. Procedures to Test Individual Difference Manipulations—We used the same training procedures described in Section 2.2 for all models, and manipulated parameters one at a time (i.e., keeping all other parameters at their defaults) from the beginning of training, since these parameters model a variety of genetic differences (Friedman et al., 2008). Performance was always assessed during the non-learning test period of 10 epochs that followed this standardized training. Accuracy remained high in all models; we were primarily interested in RTs for correct trials, as these are the typical variables used to assess individual differences in these tasks.

We ran 25 models per parameter value tested, and each run within a task used a different random seed for setting initial weights (though the same seeds were used across tasks and parameter settings to reduce noise in results). In selecting the parameter values to use, we used a range around the default and equated the range across manipulations as much as possible. The following parameter values were used: PFC to hidden connection strength (.1, .25, .5, .75, 1; note that a PFC to hidden connection strength of 0 would be equivalent to no communication between PFC and posterior cortex, so we used a low value of .1 instead for the lower bound); PFC recurrent connection strength (0, .25, .5, .75, 1); clearing on gating (0, .25, .5, .75, 1); and PFC gain (35, 50, 75, 100, 150, 200, 300). The gain manipulation is the gamma of the sigmoid rate-code activation function (O'Reilly et al., 2012) and has most of its effect on high activation values, but signal-to-noise ratio also involves inhibition of low values. Thus, when manipulating gain, we also varied the "nvar" parameter, which is the variance of the Gaussian noise kernel that is convolved with the activation function; this parameter determines the curvature of the activation function near the threshold, with larger values allowing more graded responses at the threshold. The values of nvar were set to 1.2/gain, with a minimum of .005 to keep this value from being too close to zero at high gains.

3.2.2. Results For Parameters Hypothesized to Relate to Common EF Variance

3.2.2.1. Connection Strength From PFC to Posterior (Hidden) Layer: We hypothesized that if top-down biasing from PFC is crucial for all EF tasks, the connection strength between PFC and the hidden layers should affect both the switch and Stroop tasks. As expected, both the switch cost, $F(4,120) = 4.66$, $p = .002$, $R^2 = .13$ (linear $F(1,120) = 11.51$, $p = .001$, $R^2 = .08$; quadratic $F(1,120) = 6.43$, $p = .013$, $R^2 = .05$), and Stroop interference, $F(4,120) = 3.02$, $p = .020$, $R^2 = .09$ (linear $F(1,120) = 7.27$, $p = .008$, $R^2 = .06$; cubic $F(1,120) = 4.20$, $p = .043$, $R^2 = .03$), were significantly reduced by greater connection strength from PFC to the hidden layers that perform stimulus-response mappings (see Figure 4A). The actual slopes are relatively shallow due to the fact that these scores are difference scores based on conditions that all showed significant linear decreases in RTs with increasing connection strength, all $F(1,120) > 190.73$, $p < .001$, in addition to nonlinear effects, as shown in Figure 4B and 4C. This pattern might be expected given that both trial types (e.g., switch and repeat) were mixed in blocks and so required some task set maintenance and top-down control even on easier conditions (e.g., repeat trials or congruent trials). This overall trend is consistent with the fact that, in our studies, RTs in all conditions are faster for individuals high on the Common EF factor (unpublished results). Nevertheless, the presence of significant effects on the difference scores indicates that the slope of the effect was significantly steeper for the more difficult conditions (switch trials in Color-Shape and incongruent trials in Stroop).

3.2.2.2. Signal to Noise in PFC: The second Common EF manipulation was unit gain in PFC. This parameter determines signal-to-noise ratio in the layer, which is hypothesized to be affected by tonic extracellular dopamine levels in PFC (Servan-Schreiber, Printz, & Cohen, 1990). In our models gain was yoked to the closely related parameter nvar, as described earlier. This manipulation is of particular interest given that the frequently studied COMT val$^{108/158}$met polymorphism is hypothesized to influence extracellular dopamine levels in frontal cortex, thus affecting signal-to-noise ratio, with the val allele leading to lower levels (e.g., Tunbridge et al., 2006). The effect of a higher signal-to-noise ratio in PFC is deeper attractor states, which leads to crisper representations of goals, but also stronger activations that then must be overcome when goals switch; thus, we expected this manipulation to have opposite effects on Stroop and switch performance. Signal-to-noise (i.e., gain) significantly influenced switch costs, $F(6,168) = 12.44$, $p < .001$, $R^2 = .31$ (linear $F(1,168) = 23.76$, $p < .001$, $R^2 = .10$; quadratic $F(1,168) = 23.38$, $p < .001$, $R^2 = .09$; cubic $F(1,168) = 18.90$, $p < .001$, $R^2 = .08$; 5th order $F(1,168) = 5.69$, $p = .018$, $R^2 = .02$), and Stroop interference, $F(6,168) = 14.68$, $p < .001$, $R^2 = .34$ (linear $F(1,168) = 74.53$, $p < .001$, $R^2 = .29$; quadratic $F(1,168) = 10.04$, $p = .002$, $R^2 = .04$). However, as shown in Figure 5A, the two costs showed opposite patterns of results, with higher gain resulting in higher switch costs but lower Stroop interference costs. This opposite pattern for the switch task was not due to the manipulation having a qualitatively different effect on the individual trials; as shown in Figure 5B and 5C, the RTs in both tasks both showed significant linear decreases in response to gain, all $F(1,168) > 137.76$, $p < .001$, along with nonlinear effects. However, the significantly higher switch costs with higher gain is due to the lower effect on switch trials than repeat trials. In contrast, for the Stroop task, incongruent trials benefitted more from higher gain, resulting in lower interference costs.

3.2.3. Parameters Hypothesized to Relate to Shifting-Specific Variance

3.2.3.1. Recurrent Connectivity in PFC: Interference occurs when no-longer-relevant task goals stick around to interfere with current goals. Therefore, any factors contributing to longevity of PFC representations that cannot be switched off by the BG should contribute to switch-specific effects. We hypothesized that one mechanism that might lead to such uncontrollable persistence or “stickiness” might be recurrent connectivity in the PFC, because such connectivity serves to enable representations to self reinforce and thus remain active in the absence of external input. Such stickiness may be particularly detrimental for switch tasks because the goals are rapidly changing. Higher recurrent connectivity leads to continued activation of previous goals/task sets, and that would cause interference on switch trials.

This hypothesis was confirmed by our manipulation of recurrent connection strength. As shown in Figure 6A, Color-Shape switch costs showed a significant increase as recurrent connection strength increased, $F(4,120) = 6.89$, $p < .001$, $R^2 = .19$ (linear $F(1,120) = 21.82$, $p < .001$, $R^2 = .15$), whereas Stroop interference costs showed a small decrease, $F(4,120) = 4.32$, $p = .003$, $R^2 = .13$ (linear $F(1,120) = 4.36$, $p = .039$, $R^2 = .03$; quadratic $F(1,120) = 10.08$, $p = .002$, $R^2 = .07$). An examination of the RTs in the individual conditions (Figure 6B and 6C) indicated that, though there were significant nonlinear effects on both conditions of the Color-Shape task, switch RTs showed a significant linear increase, $F(1,120) = 18.48$, $p < .001$, $R^2 = .12$, whereas repeat RTs did not, $F(1,120) = 0.27$, $p = .606$, $R^2 = .00$. This pattern resulted in larger switch costs at high levels of recurrent connectivity. In contrast, Stroop RTs did not show significant linear decreases in any condition, all $F(1,120) < 2.71$, $p > .102$, although there were significant quadratic effects on both the incongruent and neutral conditions.

3.2.3.2. Clearing on Gating: The second manipulation of interference from irrelevant goals was in a parameter affecting the degree to which representations "decay" once active maintenance is toggled off. This parameter, called clearing on gating, is the fraction of unit activity that remains after a "go" action by the BG (a decision, in essence, that it is time to either replace or use the given working memory item). Thus, a value of .25 for clearing on gating means that each unit in PFC is reduced by 25% of its previous activity on the trial after the "go" event, corresponding to relatively weak inhibitory activity, while a value of 1.0 means that the activity, and so the representations, are eliminated entirely. We used this parameter as a blanket manipulation of the varied neural mechanisms, both passive and active, that might contribute to clearing of working memory representations after they are no longer needed. The strength of GABAergic inhibition in cortex triggered by a wave of excitatory activity when a thalamocortical circuit is activated is one such variable (see, e.g., Rigas & Castro-Alamancos, 2007), but others mechanisms may contribute to this effect as well.

Our hypothesis was that higher levels of clearing on gating would lead to lower switch costs, but should not affect tasks in which switching is not important. This prediction was confirmed: As shown in Figure 7A, Color-Shape switch costs showed a significant increase as clearing on gating decreased, $F(4,120) = 15.78$, $p < .001$, $R^2 = .34$ (linear $F(1,120) = 57.24$, $p < .001$, $R^2 = .31$), whereas Stroop interference costs were not significantly affected by clearing on gating, $F(4,120) = 1.64$, $p = .169$, $R^2 = .05$. This effect on switch costs was driven by the switch cost for cue naming RTs, $F(4,120) = 26.46$, $p < .001$, $R^2 = .47$; the effect on switch costs for task performance RTs was not significant, $F(4,120) = 0.99$, $p = .417$, $R^2 = .03$. An examination of the RTs in the individual conditions (see Figure 7B) indicated that Color-Shape switch RTs showed a significant linear decrease in RT with lower clearing on gating, $F(1,120) = 16.68$, $p < .001$, $R^2 = .12$, as did the repeat RTs, $F(1,120) = 57.91$, $p < .001$, $R^2 = .32$ (quadratic also significant), but repeat RTs decreased more than switch RTs, consistent with the idea that less clearing on gating resulted in more persistent task sets that helped speed re-establishment on repeat trials.

The fact that lower clearing also, on average, speeded responses on switch trials is counterintuitive. Because much of the time to establish the new task set is taken by developing activations, switching from an old to new task set representation is actually faster than developing a new representation from scratch. The old representation in effect provides a small "jump-start" for the new representation. Stroop RTs did not show significant decreases with lower clearing on gating, all linear $F(1,120) < 2.20$, $p > .072$, though there was a significant quadratic effect on neutral RTs (see Figure 7C). Because the Stroop RTs did not include cue naming times, they did not benefit from a "jump start" from previous trials. However, were such naming times to be included, we would expect a similar jump start for all trial types (since all involve the same goal and no switching), leading to no effect on the difference scores.

4. Discussion

We presented a biologically based neural network model of task switching that showed individual differences in two separable components of switch costs — Common EF and

Shifting-Specific (summarized in Figure 2) — in addition to reproducing many benchmark findings from the task-switching literature. Specifically, to simulate individual differences we manipulated four parameters to affect gated active maintenance of task goals and top-down biasing (Common EF), and residual activation of no-longer-relevant goals or goal "stickiness" (Shifting-Specific). Whereas the top-down biasing manipulation (PFC to posterior (hidden) layers connection strength) had beneficial influences on both the inhibition and switch tasks, the strength of goal maintenance (unit gain in PFC, determining signal-to-noise ratio) had opposite effects on the two tasks. These results are consistent with a trade-off between stability and flexibility (Goschke, 2000), influenced by individual differences in cortical signal-to-noise ratio that is modulated by (among other things) dopamine function (Colzato, Waszak, Nieuwenhuis, Posthuma, & Hommel, 2010). The manipulations of non-gated maintenance (stickiness) in PFC had significant overall effects on switch costs, while having either small opposite (recurrent connection strength) or null (clearing on gating) effects on Stroop interference, as predicted.

4.1. Mapping EF Components to Biological Mechanisms

These results demonstrate that detailed neural network models based closely on known biology can be useful in distinguishing the neural mechanisms underlying different components of individual differences in EF. Mapping from biological differences to their effects on composite measures of comparative performance across many tasks is a complex undertaking. This type of theory development may thus particularly stand to benefit from the application of detailed computational models that track the interacting effects of many theoretical commitments, and can reveal unexpected consequences of manipulations. For instance, we did not fully predict the effects of either PFC gain or recurrent connectivity on both switch and Stroop tasks, so the model directly informed our perspective on the likely contributions of such factors to particular EF component abilities.

More specifically, our model suggests some mechanistic underpinnings of the individual differences in different components of EF abilities. Parameter manipulations in the model (and therefore, we predict, genetic and learned differences in brains) that contribute to the uncontrolled (un-gated), automatic persistence of representations are good candidates for the Shifting-Specific component of EF identified by Friedman et al. (2008). Such neural factors include the density and strength of recurrent connections within PFC, intrinsic tendency of neurons to fire persistently (hysteresis), and signal-to-noise ratio (gain) as influenced by dopamine and the COMT genetic variation, among others. This component should also be controlled by neural factors that affect active clearing of PFC representations with gating (as captured by the clearing on gating parameter). These factors would be complex, and are less well understood, since there are no known inhibitory neurons whose function is unique to gating interactions. Any factors causing cortical inhibitory neurons to fire strongly specifically when a wave of activity is first gated through the thalamus should also affect the Shifting-Specific components; however, we are not aware of any such properties that are currently known.

The Common EF component should be determined by variance in the quality/strength of gated PFC representations and their effect on processing in the rest of the cognitive system.

Such mechanisms include the influence of PFC on other brain areas, including density and strength of connections. They also include the variety of mechanisms determining how strongly the BG affect cortical processing, including strength of BG connections with thalamus, and those from thalamus to PFC, etc. Importantly, the general tendency for the BG to gate should also affect the Common EF component, in the sense that incorrectly gating might lead to goal neglect or performing the wrong task. Our model could not capture this effect, as its limited training regime consisting of only two tasks trained gating so effectively that gating was effectively perfect. This is in sharp contrast to everyday human experience, in which it is important to not attend to and maintain each new experience. Variances in this tendency can be captured in the PBWM framework, and have been used as the basis of a theory that, among other variables, decreased tonic dopamine levels in the BG lead to gating too rarely, causing task neglect and, in extreme cases, attentional deficit disorder (Frank et al, 2007).

Importantly, these simulations provide a mechanistic account for the surprising tradeoffs observed in many studies between switching and other components of EF. In particular, the PFC gain parameter and, to a smaller extent, the recurrent connection parameter, decreased Stroop interference but increased switch costs, mimicking the pattern of correlations seen with certain variables (Miyake & Friedman, 2012; see also Blackwell et al., this issue). These manipulations suggest that one source of these tradeoffs is in the strength of goal maintenance (tied to dopamine function in cortex, among other biological factors), versus stickiness of representations (affected by the clearing on gating parameter and hypothesized to be related to GABA), and as such provide testable predictions. These results are compatible with a recent finding that having more val alleles in the COMT val$^{108/158}$met polymorphism, which results in lower tonic DA in PFC, is associated with lower switch costs (Colzato et al., 2010), despite being also associated with worse executive performance in tasks such as *n*-back and the Wisconsin Card Sorting Test (e.g., Egan et al., 2001; Caldú et al., 2007). In our model the parameter most related to the COMT polymorphism would be gain in PFC units (hypothesized to affect the Common EF component), and it shows the same trade-off in terms of Stroop and switch.

In discussing some potential mechanistic underpinnings of dissociable EF abilities, it is worth clarifying that the mapping is complex and not one-to-one. Each biological mechanism is likely to contribute to performance in many EF tasks, but to different degrees; as such, the loading of each mechanism on each factor will depend upon the analysis used to separate components (e.g., whether components are constrained to be orthogonal, and whether a single common component is included). In addition to the complexity of single biological factors contributing different amounts of variance to multiple EF components (as in the effects of gain and recurrent connection strength), it is important to recognize that our consideration of the Common EF and Shifting-Specific components in no way suggests that there are precisely two underlying biological factors, or even two closely functionally related classes of such factors. Instead, each component can be affected by biological variations with many different computational functions. For example, in this model, the Shifting-Specific factor seems to be affected not only by uncontrollable persistence of PFC representations, as captured by our recurrent connectivity manipulation, but also by the strength of mechanisms which specifically clear or disrupt those representations when gating signals the need for them to change, as captured by our clearing-on-gating mechanism. While these are only two classes of functional mechanisms, it is likely that there are other classes we have not identified.

In a similar vein, we have noted in previous work that while response inhibition, working memory updating, and task shifting are some of the most commonly discussed EFs, there are other possible separable EFs (e.g., Friedman & Miyake, 2004; Friedman et al., 2008; Miyake et al., 2000). Thus, it is important to note here that we do not assume there are precisely three components of EF, but merely that different EFs tasks tap both common and distinct mechanisms (the Unity/Diversity model; Miyake & Friedman, 2012). While existing studies are insufficient to identify all components of executive control across all possible types of tasks, they are sufficient to demonstrate that there is a switching factor that is distinct from a common factor, and it is that distinction we address.

4.2. Extensions and Future Directions

Although the current model is rich enough to produce hypotheses regarding several important mechanisms underlying different components of EF ability, it could be productively expanded in future work to provide a more complete model of task switching, which would in turn make more specific predictions about that mapping. The current model begins to address meta-control by capturing some of the mechanisms controlling switching between different cognitive control representations, but two expansions on different aspects of meta-control suggest themselves in particular.

The first such expansion on mechanisms of meta-control is an adaptive feedback mechanism, as in work addressing anterior cingulate cortex (ACC) contributions to EF. Although our current understanding of the relevant systems is incomplete (for instance, the role of brainstem neuromodulatory systems in performance-based adaptation of cognitive control, Krichmar, 2008), adding even a simple mechanism that increases the strength of maintained goals in response to poor task performance has been shown to improve model fit to inter-trial dependencies in task performance (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Brown et al., 2007). Although such data has not, to our knowledge, been analyzed in terms of individual differences, much less a multi-factor model, inclusion of such mechanisms would improve the fidelity of the model, and so likely provide suggestions of other mechanisms that might contribute to the EF components we address here. For instance, mechanistic (biological) differences contributing to stronger error-driven goal activation could be hypothesized to contribute to the Common EF component, since that mechanism would contribute to success in all tasks. However, these same mechanisms might also contribute negatively to the Shifting-Specific component, if that feedback enhances activation of the current goal after an error (or conflict), and thereby makes that goal more persistent and so more difficult to replace with the next goal on switch trials. Future modeling work would help establish how specific individual differences in mechanisms underlying this trial-to-trial task set regulation map to specific EF components.

Another extension to the model that would address a second proposed explanation for asymmetrical switch costs is persistent inhibition of previous task sets: If the dominant task must be inhibited more strongly, persistent inhibition would make it harder to re-activate (Koch, Gade, Schuch & Philipp, 2010). Support for this explanation comes from studies showing that switch costs are higher when switching to a task that has been recently switched away from (the "backward inhibition" effect; e.g., Mayr & Keele, 2000). We regard this explanation as incomplete, since it requires some further mechanism whereby a neural representation becomes harder to re-activate after it has been inhibited, as the firing of inhibitory neurons is not persistent across the relevant time scales of several seconds. The competitive learning effects hypothesized and supported by Norman and colleagues offers one such explanation; they propose that neurons that are briefly activated, then are competitively inhibited, weaken their connections (Norman, Newman, & Detre, 2007). This learning rule is not employed in this particular model, but is fully consistent with it. Another possible explanation is a learned connection from an incorrect task set to the subthalamic nucleus (STN), which appears to delay or halt actions (e.g., Frank, 2006). This connection could be strengthened whenever a task set is not properly reconfigured, and so produces a wrong answer. This persistent learning would cause a slower RT, but the time course of synaptic potentiation does not, by itself, explain the strong difference in recently used versus more distant task sets.

Another significant area for enhancement of the model is in including an "outer loop" for meta-task goal maintenance. The current model switches only between the two trained tasks, and so has goal representations for only those tasks. As such, it cannot display goal neglect (e.g., Duncan, Emslie, Williams, Johnson, & Freer, 1996) by switching to a goal irrelevant to the current task. In contrast, humans have trained on a huge variety of "tasks" over the lifespan instead of just two, and have goals outside of the current task. Our model does not show much residual switch costs with long preparation times (the switch costs for actual performance trials are small or nonexistent depending on parameters), which we take to be consistent with the hypothesis that infrequent failure to prepare the current task set is one prominent source of residual switch costs (De Jong, 2000). Such "failure to engage" could in part reflect goal neglect at the meta-control level — i.e., failing to perform the more abstract task of attending to and using the task cues to switch goal representations.

Humans may use such "outer loop" goals, like performing task switching, by hierarchical active maintenance of sub-goals and superordinate-goals simultaneously in PFC (e.g., Badre & D'Esposito, 2007; Koechlin, Ody & Kouneieher, 2003). These meta-task goals would be represented in the same type of PFC/BG circuit, but in different areas of PFC. In the current model, this "task set" is represented only implicitly, in the switching rules learned by PFC and BG in response to a task cue; there is no "switching" goal held in active maintenance. If such meta-task goals were actively maintained, they would be consistent across sub-tasks. They would not need frequent switching, and so would not rely on switching-specific mechanisms; rather they would rely more upon the parameters controlling strength of maintenance, which we have here identified with the Common EF component. As such, this aspect of the model might be more accurate for very well practiced task switches, for which learning can transfer knowledge of switch rules from active maintenance or episodic memory to cortical synaptic memory as it has in the model. Evidence for this explicit structure of maintained goals in novel tasks is mixed (Crittenden & Duncan, 2014; Reynolds, O'Reilly, Cohen & Braver, 2012), perhaps because many factors may influence whether goals (whether sub- or superordinate) are maintained, versus retrieved from episodic memory (e.g., Braver, Gray & Burgess, 2007; Chatham, Frank & Munakata, 2009; D'Ardenne et al., 2012). For example, task switching might be more likely to rely on the maintenance of both sub- and superordinate control representations, depending on the time pressure and frequency of switching in the paradigm used. By contrast, retrieval from episodic memory might be a more prominent feature of tasks requiring less speeded responses, or less frequent access to meta-control representations. Further empirical and modeling work will be needed to identify how and when these mechanisms contribute to meta-control of cognition.

Adding "meta-task-set" and performance-based adaptive mechanisms to the model are only two examples of how such models might be made more complete and accurate, and so produce better hypotheses on the brain mechanisms underlying human EF. Given the wealth of data pertinent to the issue, further work using explicit, implemented models that can span levels of analysis could be particularly useful in this domain.

4.3. Summary and Conclusion

The current model extends the literature on computational models of task switching to include a model that can account for individual differences in switch costs. Because the model built on the existing PBWM framework (Frank et al., 2007), it was integrated with other EF tasks, and we used this integration to compare how parameter manipulations influenced tasks tapping different components of EF (i.e., Common EF vs. Shifting-Specific; Miyake & Friedman, 2012).

We confirmed two predictions about the underlying mechanisms for these factors within the model. First, parameters that increased goal "stickiness" increased switch costs but not Stroop interference, consistent with our prediction that the Shifting-Specific component of the Unity/Diversity model may reflect the extent to which no-longer-relevant goals persist in PFC. Second, parameters that affected the model's ability to actively maintain goals in the PFC layer and use those goals to bias processing in other brain areas influenced both the switch and Stroop tasks, consistent with our hypothesis that this ability is key to Common EF variation. Importantly, one of the manipulations (PFC unit gain, affecting signal-to-noise ratio in PFC representations) showed opposite effects on Stroop interference and switch costs, mirroring stability-flexibility trade-offs evident in individual difference (see Miyake & Friedman, 2012) and molecular genetic literature (e.g., Colzato et al., 2010). This model thus provides a foundation for future investigation into the biological and genetic mechanisms underlying variation in executive control.

Acknowledgments

This research was supported by grants MH063207 and MH079485 from the National Institute of Health. We thank Akira Miyake and Trent Kriete for helpful comments and suggestions on this project, as well as all the members of the Interdisciplinary Behavioral Science Center on the Determinants of Executive Function and Dysfunction (NIMH P50 MH079485) for their discussions of and contributions to the ideas presented here. We also thank Iring Koch for a particularly thorough and helpful signed review that improved the manuscript in many ways.

References

Allport, DA.; Styles, EA.; Hsieh, S. Shifting intentional set: Exploring the dynamic control of tasks. In: Umilta, C.; Moscovitch, M., editors. Attention and Performance XV. Cambridge, MA: MIT Press; 1994. p. 421-452.

Altamirano LJ, Miyake A, Whitmer AJ. When mental inflexibility facilitates executive control: Beneficial side effects of ruminative tendencies on goal maintenance. Psychological Science. 2010; 21:1377–1382. [PubMed: 20798398]

Altmann EM, Gray WD. An integrated model of cognitive control in task switching. Psychological Review. 2008; 115:602–639. [PubMed: 18729594]

Arrington CM, Logan GD, Schneider DW. Separating cue encoding from target processing in the explicit task-cuing procedure: Are there “true” task switch effects? Journal of Experimental Psychology: Learning, Memory, and Cognition. 2007; 33:484–502.

Badre D, D’Esposito M. Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. Journal of Cognitive Neuroscience. 2007; 19:2082–2099. [PubMed: 17892391]

Bilder RM. Executive control: balancing stability and flexibility via the duality of evolutionary neuroanatomical trends. Dialogues In Clinical Neuroscience. 2012; 14:39–47. [PubMed: 22577303]

Blackwell, KA.; Chatham, CH.; Wiseheart, M.; Munakata, Y. A developmental window into trade-offs in executive function: The case of task switching versus response inhibition in 6-year-olds. (this issue)Manuscript under review

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. Conflict monitoring and cognitive control. Psychological Review. 2001; 108:624–652. [PubMed: 11488380]

Braver, TS.; Gray, JR.; Burgess, GC. Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In: Conway, ARA.; Jarrold, C.; Kane, MJ.; Miyake, A.; Towse, JN., editors. Variation in working memory. New York: Oxford University Press; 2007. p. 76-106.

Brown JW, Reynolds JR, Braver TS. A computational neural model of fractionated conflict-control mechanisms in task-switching. Cognitive Psychology. 2007; 55:37–85. [PubMed: 17078941]

Caldú X, Vendrell P, Bartres-Faz D, Clemente I, Bargallo N, Jurado MA, Serra-Grabulosa JM, Junqué C. Impact of the COMT Val108/158 Met and DAT genotypes on prefrontal function in healthy subjects. Neuroimage. 2007; 37:1437–1444. [PubMed: 17689985]

Chatham CH, Claus ED, Banich MT, Curran T, Kim A, Munakata Y. Cognitive control reflects context monitoring, not motoric stopping, in response inhibition. PLoS One. 2012; 7:e31546. [PubMed: 22384038]

Chatham CH, Frank MJ, Munakata Y. Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. Proceedings of the National Academy of Sciences. 2009; 106:5529–5533.

Chatham CH, Herd SA, Brant AM, Hazy TE, Miyake A, O’Reilly R, Friedman NP. From an executive network to executive control: a computational model of the n-back task. Journal of Cognitive Neuroscience. 2011; 23:3598–3619. [PubMed: 21563882]

Cohen JD, Dunbar K, McClelland JL. On the control of automatic processes: A parallel distributed processing model of the Stroop effect. Psychological Review. 1990; 97:332–361. [PubMed: 2200075]

Collette F, Van der Linden M, Laureys S, Delfiore G, Degueldre C, Luxen A, Salmon E. Exploring the unity and diversity of the neural substrates of executive functioning. Human Brain Mapping. 2005; 25:409–423. [PubMed: 15852470]

Collins AGE, Frank MJ. Cognitive control over learning: Creating, clustering, and generalizing task-set structure. Psychological Review. 2013; 120:190–229. [PubMed: 23356780]

Colzato LS, Waszak F, Nieuwenhuis S, Posthuma D, Hommel B. The flexible mind is associated with the catechol-o-methyltransferase (COMT) Val^{158}Met polymorphism: Evidence for a role of dopamine in the control of task-switching. Neuropsychologia. 2010; 48:2764–2768. [PubMed: 20434465]

Cools, R. Chemical neuromodulation of goal-directed behavior. In: Todd, PM.; Hills, TT.; Robbins, TW.; Lupp, J., editors. Cognitive search: Evolution, algorithms, and the brain. Vol. 9. Cambridge, MA: MIT Press; 2012. p. 111-123.Strüngmann Forum Report

Crittenden BM, Duncan J. Task difficulty manipulation reveals multiple demand activity but no frontal lobe hierarchy. Cerebral Cortex. 2014; 24:532–540. [PubMed: 23131804]

Curtis CE, Lee D. Beyond working memory: The role of persistent activity in decision making. Trends in Cognitive Sciences. 2011; 14:216–222. [PubMed: 20381406]

D'Ardenne K, Eshel N, Luka J, Lenartowicz A, Nystrom LE, Cohen JD. Role of prefrontal cortex and the midbrain dopamine system in working memory updating. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:19900–19909. [PubMed: 23086162]

De Jong, R. An intention-activation account of residual switch costs. In: Monsell, S.; Driver, J., editors. Control of cognitive processes: Attention and performance XVIII. Cambridge, MA: MIT Press; 2000. p. 357-376.

Deco G, Rolls ET. Synaptic and spiking dynamics underlying reward reversal in the orbitofrontal cortex. Cerebral Cortex. 2005; 15:15–30. [PubMed: 15238449]

Desimone R, Duncan J. Neural mechanisms of selective visual attention. Annual Review of Neuroscience. 1995; 18:193–222.

Duncan J, Emslie H, Williams P, Johnson R, Freer C. Intelligence and the frontal lobe: The organization of goal-directed behavior. Cognitive psychology. 1996; 30:257–303. [PubMed: 8660786]

Egan MF, Goldberg TE, Kolachana BS, Callicott JH, Mazzanti CM, Straub RE, Goldman D, Weinberger DR. Effect of COMT val$^{108/158}$met genotype on frontal lobe function and risk for schizophrenia. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98:6917–6922. [PubMed: 11381111]

Engle, RW.; Kane, M.; Tuholski, S. Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In: Miyake, A.; Shah, P., editors. Models of working memory: Mechanisms of active maintenance and executive control. New York: Cambridge University Press; 1999. p. 102-134.

Frank MJ. Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. Neural Networks. 2006; 19:1120–1136. [PubMed: 16945502]

Frank MJ, Loughry B, O'Reilly RC. Interactions between the frontal cortex and basal ganglia in working memory: A computational model. Cognitive, Affective, and Behavioral Neuroscience. 2001; 1:137–160.

Frank MJ, Seeberger LC, O'Reilly RC. By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. Science. 2004; 306:1940–1943. [PubMed: 15528409]

Friedman, NP.; Herd, SA.; Brant, AM.; Hazy, TE.; Kriete, T.; Chatham, CH.; O'Reilly, RC. Neural network models of individual differences in executive functions. 2014. Manuscript in preparation

Friedman N, Miyake A, Corley R, Young S, DeFries J, Hewitt J. Not all executive functions are related to intelligence. Psychological Science. 2006; 17:172–179. [PubMed: 16466426]

Friedman NP, Haberstick BC, Willcutt EG, Miyake A, Young SE, Corley RP, Hewitt JK. Greater attention problems during childhood predict poorer executive functioning in late adolescence. Psychological science. 2007; 18:893–900. [PubMed: 17894607]

Friedman NP, Miyake A. The relations among inhibition and interference control functions: A latent-variable analysis. Journal of Experimental Psychology: General. 2004; 133:101–135. [PubMed: 14979754]

Friedman NP, Miyake A, Robinson JL, Hewitt JK. Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. Developmental Psychology. 2011; 47:1410–1430. [PubMed: 21668099]

Friedman NP, Miyake A, Young SE, Defries JC, Corley RP, Hewitt JK. Individual differences in executive functions are almost entirely genetic in origin. Journal of Experimental Psychology: General. 2008; 137:201–225. [PubMed: 18473654]

Goschke, T. Intentional reconfiguration and involuntary persistence in task set switching. In: Monsell, S.; Driver, J., editors. Control of cognitive processes: Attention and performance XVIII. Cambridge, MA: MIT Press; 2000. p. 331-355.

Hazy TE, Frank MJ, O'Reilly RC. Banishing the homunculus: Making working memory work. Neuroscience. 2006; 139:105–118. [PubMed: 16343792]

Hazy TE, Frank MJ, O'Reilly RC. Towards an executive without a homunculus: Computational models of the prefrontal cortex/basal ganglia system. Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences. 2007; 362:105–118.

Hazy TE, Frank MJ, O'Reilly RC. Neural mechanisms of acquired phasic dopamine responses in learning. Neuroscience and Biobehavioral Reviews. 2010; 34:701–720. [PubMed: 19944716]

Hedden T, Yoon C. Individual differences in executive processing predict susceptibility to interference in verbal working memory. Neuropsychology. 2006; 20:511–528. [PubMed: 16938014]

Herd SA, Banich MT, O'Reilly RC. Neural mechanisms of cognitive control: An integrative model of Stroop task performance and fMRI data. Journal of Cognitive Neuroscience. 2006; 18:22–32. [PubMed: 16417680]

Kane MJ, Engle RW. Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. Journal of Experimental Psychology: General. 2003; 132:47–70. [PubMed: 12656297]

Kiesel A, Steinhauser M, Wendt M, Falkenstein M, Jost K, Philipp AM, Koch I. Control and interference in task switching — A review. Psychological Bulletin. 2010; 136:849–874. [PubMed: 20804238]

Koch I, Gade M, Schuch S, Philipp AM. The role of inhibition in task switching: A review. Psychonomic Bulletin & Review. 2010; 17:1–14. [PubMed: 20081154]

Koechlin E, Ody C, Kouneiher F. Neuroscience: The architecture of cognitive control in the human prefrontal cortex. Science. 2003; 302:1181–1184. [PubMed: 14615530]

Krichmar JA. The Neuromodulatory system: A framework for survival and adaptive behavior in a challenging world. Adaptive Behavior. 2008; 16:385–399.

Kray J, Lindenberger U. Adult age differences in task switching. Psychology and Aging. 2000; 15:126–147. [PubMed: 10755295]

Krueger KA, Dayan P. Flexible shaping: How learning in small steps helps. Cognition. 2009; 110:380–394. [PubMed: 19121518]

Lehto JE, Juujarvi P, Kooistra L, Pulkkinen L. Dimensions of executive functioning: Evidence from children. British Journal of Developmental Psychology. 2003; 21:59–80.

Logan GD, Bundesen C. Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? Journal of Experimental Psychology: Human Perception and Performance. 2003; 29:575–599. [PubMed: 12848327]

Logan GD, Gordon RD. Executive control of visual attention in dual-task situations. Psychological Review. 2001; 108:393–434. [PubMed: 11381835]

Logan GD, Schneider DW. Distinguishing reconfiguration and compound-cue retrieval in task switching. Psychologica Belgica. 2010; 50:413–433.

Mayr U, Keele SW. Changing internal constraints on action: the role of backward inhibition. Journal of Experimental Psychology: General. 2000; 129:4–26. [PubMed: 10756484]

Meiran N, Kessler Y, Adi-Japha E. Control by action representation and input selection (CARIS): A theoretical framework for task switching. Psychological Research. 2008; 72:473–500. [PubMed: 18350316]

Miller EK. The prefrontal cortex and cognitive control. Nature Reviews Neuroscience. 2000; 1:59–65.

Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annual Review of Neuroscience. 2001; 24:167–202.

Miyake A, Emerson MJ, Padilla F, Ahn J. Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. Acta Psychologica. 2004; 115:123–142. [PubMed: 14962397]

Miyake A, Friedman NP. The nature and organization of individual differences in executive functions: Four general conclusions. Current Directions in Psychological Science. 2012; 21:8–14. [PubMed: 22773897]

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. Cognitive Psychology. 2000; 41:49–100. [PubMed: 10945922]

Monsell S. Task switching. Trends in Cognitive Sciences. 2003; 7:134–140. [PubMed: 12639695]

Morton JB, Munakata Y. Active versus latent representations: A neural network model of perseveration, dissociation, and decalage. Developmental Psychobiology. 2002; 40:255–265. [PubMed: 11891637]

Munakata Y, Herd SA, Chatham CH, Depue BE, Banich MT, O’Reilly RC. A unified framework for inhibitory control. Trends in Cognitive Sciences. 2011; 15:453–459. [PubMed: 21889391]

Norman KA, Newman EL, Detre G. A neural network model of retrieval-induced forgetting. Psychological Review. 2007; 114:887–953. [PubMed: 17907868]

Newell, A. Unified Theories of Cognition. Cambridge, MA: Harvard University Press; 1990.

O’Reilly R. Biologically based computational models of high-level cognition. Science. 2006; 314:91–94. [PubMed: 17023651]

O’Reilly, RC.; Braver, TS.; Cohen, JD. A biologically based computational model of working memory. In: Miyake, A.; Shah, P., editors. Models of working memory: Mechanisms of active maintenance and executive control. New York: Cambridge University Press; 1999. p. 375-411.

O’Reilly RC, Frank MJ. Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. Neural Computation. 2006; 18:283–328. [PubMed: 16378516]

O’Reilly RC, Frank MJ, Hazy TE, Watz B. PVLV: The primary value and learned value Pavlovian learning algorithm. Behavioral Neuroscience. 2007; 121:31–49. [PubMed: 17324049]

O’Reilly, RC.; Hazy, TE.; Herd, SA. The Leabra cognitive architecture: How to play 20 principles with nature and win!. In: Chipman, S., editor. The Oxford handbook of cognitive science. New York: Oxford University Press; (in press)

O’Reilly, RC.; Munakata, Y. Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain. Cambridge, MA: MIT Press; 2000.

O’Reilly, RC.; Munakata, Y.; Frank, MJ.; Hazy, TE., et al. Computational cognitive neuroscience. 1. Wiki Book; 2012. URL: http://ccnbook.colorado.edu

Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. Applied Psychological Measurement. 1977; 1:385–401.

Reynolds JR, Braver TS, Brown JW, Van der Stigchel S. Computational and neural mechanisms of task switching. Neurocomputing. 2006; 69:1332–1336.

Reynolds JR, O’Reilly RC, Cohen JD, Braver TS. The function and organization of lateral prefrontal cortex: a test of competing hypotheses. PloS One. 2012; 7:e30284. [PubMed: 22355309]

Rigas P, Castro-Alamancos MA. Thalamocortical up states: Differential effects of intrinsic and extrinsic cortical inputs on persistent activity. Journal of Neuroscience. 2007; 27:4261–4272. [PubMed: 17442810]

Rogers RD, Monsell S. Costs of a predictable switch between simple cognitive tasks. Journal of Experimental Psychology: General. 1995; 124:207–231.

Rougier NP, Noelle D, Braver TS, Cohen JD, O’Reilly RC. Prefrontal cortex and the flexibility of cognitive control: Rules without symbols. Proceedings of the National Academy of Sciences. 2005; 102:7338–7343.

Rubin O, Meiran N. On the origins of the task mixing cost in the cuing task-switching paradigm. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31:1477–1491.

Servan-Schreiber D, Printz H, Cohen JD. A network model of catecholamine effects: Gain, signal-to-noise ratio, and behavior. Science. 1990; 249:892–895. [PubMed: 2392679]

Skinner, BF. The Behavior of Organisms. New York: Appleton-Century-Crofts; 1938.

Sohn MH, Anderson JR. Task preparation and task repetition: Two-component model of task switching. Journal of Experimental Psychology, General. 2001; 130:764–778. [PubMed: 11757879]

Stroop JR. Studies of interference in serial verbal reactions. Journal of Experimental Psychology. 1935; 18:643–662.

Teuber HL. Unity and diversity of frontal lobe functions. Acta Neurobiologiae Experimentalis. 1972; 32:615–656. [PubMed: 4627626]

Tunbridge EM, Harrison PJ, Weinberger DR. Catechol-O-methyltransferase, cognition, and psychosis: Val^{158}Met and beyond. Biological Psychiatry. 2006; 60:141–151. [PubMed: 16476412]

Vandierendonck A, Liefooghe B, Verbruggen F. Task switching: Interplay of reconfiguration and interference control. Psychological Bulletin. 2010; 136:601–626. [PubMed: 20565170]

Wechsler, D. WAIS-III: Wechsler adult intelligence scale. San Antonio, TX: Psychological Corporation; 1997.

Young SE, Friedman NP, Miyake A, Willcutt EG, Corley RP, Haberstick BC, Hewitt JK. Behavioral disinhibition: Liability for externalizing spectrum disorders and its genetic and environmental relation to response inhibition across adolescence. Journal of Abnormal Psychology. 2009; 118:117–130. [PubMed: 19222319]

Highlights

- We present a neural network model of individual differences in task switching.
- We compared results for a Stroop model that captures general executive ability.
- Active goal maintenance and top-down biasing affected both switch and Stroop cost.
- Persistence of no-longer-relevant goals increased switch cost but not Stroop cost.
- These mechanisms may create some of the unity and diversity of executive functions.

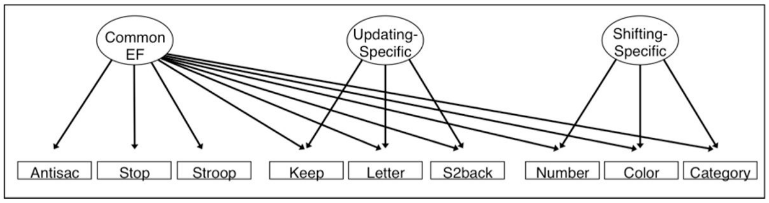

Figure 1.

Unity/Diversity model from Friedman et al. (2008). There is a Common EF latent variable on which all nine EF tasks load, as well as two “nested” latent variables on which the updating and shifting tasks, respectively, also load. The Common EF variance turned out to be isomorphic with the Inhibiting latent variable (see Friedman et al., 2008); there was not significant inhibiting-specific variance. Because the Common EF factor captures the variance common to all three EFs, the Updating-specific and Shifting-specific factors capture the variance that is unique to updating and shifting, respectively. Hence, they are uncorrelated with the Common EF factor and with each other. Antisac = antisaccade, Stop = stop-signal, Letter = letter memory, S2back = spatial 2-back, Number = number-letter, Color = color-shape, Category = category-switch.

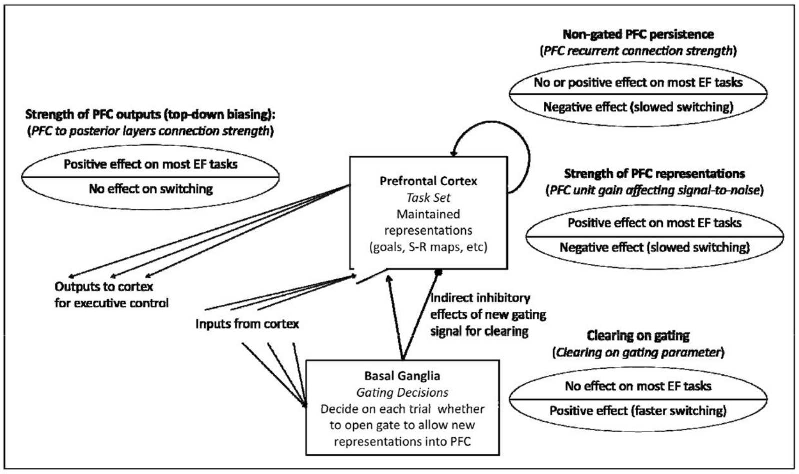
Non-gated PFC persistence
(PFC recurrent connection strength)
No or positive effect on most EF tasks
Negative effect (slowed switching)
Strength of PFC outputs (top-down biasing):
(PFC to posterior layers connection strength)
Positive effect on most EF tasks
No effect on switching
Prefrontal Cortex
Task Set
Maintained representations
(goals, S-R maps, etc)
Strength of PFC representations
(PFC unit gain affecting signal-to-noise)
Positive effect on most EF tasks
Negative effect (slowed switching)
Outputs to cortex for executive control
Inputs from cortex
Indirect inhibitory effects of new gating signal for clearing
Clearing on gating
(Clearing on gating parameter)
Basal Ganglia
Gating Decisions
Decide on each trial whether to open gate to allow new representations into PFC
No effect on most EF tasks
Positive effect (faster switching)

Figure 2.

Key model components and predictions for contributions to individual difference in EF. Hypothesized mechanisms influencing Common EF and Shifting-Specific components are depicted above the ovals (model parameter used to manipulate each mechanism in parentheses), and inside the ovals are the predicted effects on Common EF (top) and Shifting-Specific (bottom). Arrows are excitatory connections. For the gating connections, the connection “opens” the gate to allow new representations into PFC for maintenance under appropriate conditions, and the terminal circle is an indirect inhibitory effect of the basal ganglia gating signal; this wave of inhibitory cortical activity helps “clear” old PFC representations to allow new representations to come in. The key Common EF mechanisms we manipulate are Strength of PFC outputs and representations. The strength of PFC outputs (e.g, corticortical and cortico-thalamo-cortical connection density and average synaptic strength) has a positive effect on Common EF since it allows makes all task sets to have a stronger top-down biasing effect on posterior areas. It has no effect on switching speed because it does not affect the persistence of PFC representations in any way. The strength of PFC representations (i.e., signal-to-noise ratio in PFC) also has a positive effect on Common EF because it enables the task set representations to be stronger and more precise. However, it slows down switching, as stronger representations are more difficult to dislodge when no longer relevant. The key Shifting-Specific mechanisms we manipulated were non-gated persistence of PFC representations and clearing on gating. Non-gated PFC persistence (manipulated with recurrent connection strength) slows down switching because it slows down the process of replacing task sets. It generally has no effect (as in our parameter manipulation and tasks) or a positive effect on many EF tasks in which persistence of task sets is useful, or when persistence is linked to stronger PFC representations. Clearing on gating speeds up switching while not affecting most EF tasks, as it was implemented to have no effect outside of the moment of gating. Many physiological sources of clearing differences (such as strength of inhibitory interneurons in PFC) would likely produce effects on other EF components as well.

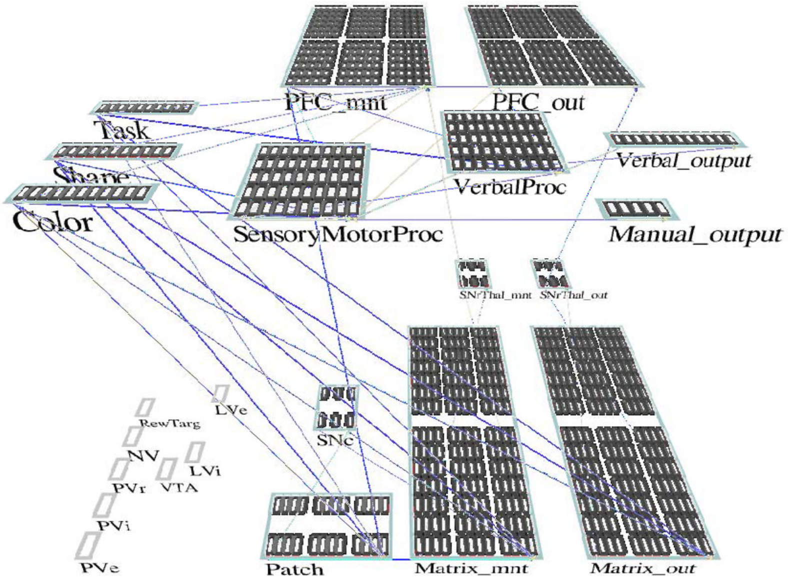

Figure 3.

Model Architecture. Each small rectangle is a single unit. Lines to the bottom right of a layer are incoming connections, while those from the left are outgoing connections. The general flow of information processing proceeds from left to right, from Color, Shape, and Task inputs, through the two Processing (hidden) layers, to the Verbal and Manual outputs. The PFC layer holds information in active states from previous trials and that information modulates processing in the two hidden layers. The PFC layer is composed of stripes, each of which is “gated” to maintain information based on the result of learning in the Matrix (striatum). This learning is in turn based on a predicted reward, calculated in the PVLV module (lower left), which is shown without units and connections to make other connections visible. The PBWM framework and PVLV system are described in detail elsewhere; see text for references.

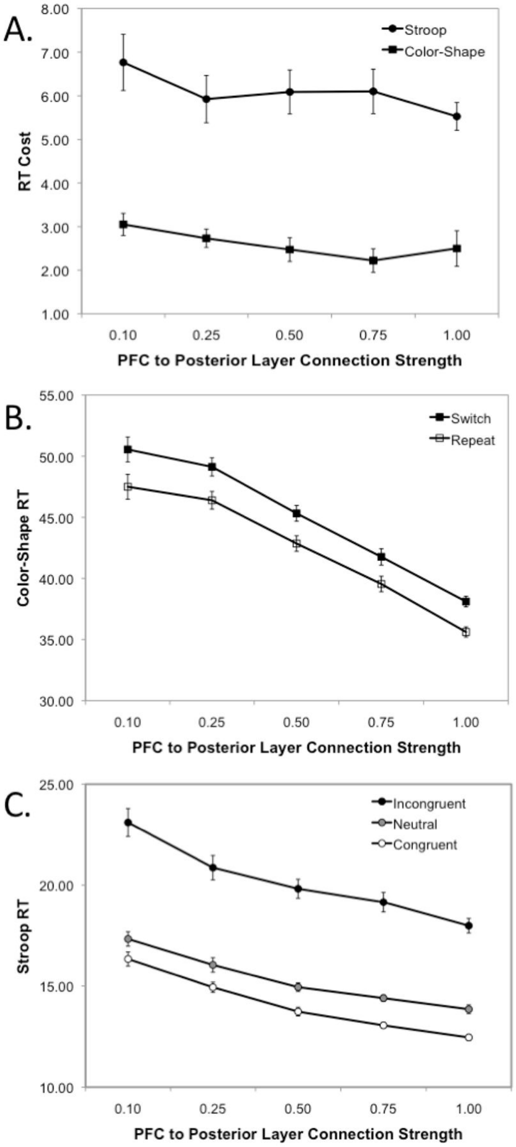

Figure 4.

Interference and Switch costs (A), and RTs in the (B) Color Shape task and (C) Stroop Task as a function of PFC to posterior (hidden) layer connection strength. Bars indicate ± 2 standard errors, approximating 95% confidence intervals.

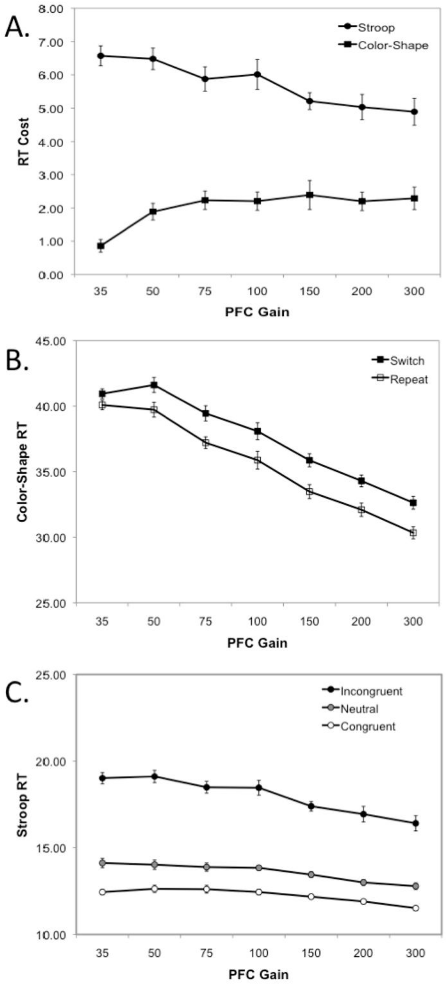

Figure 5.

Interference and Switch costs (A), and RTs in the (B) Color Shape task and (C) Stroop Task as a function of PFC gain. Bars indicate ± 2 standard errors, approximating 95% confidence intervals.

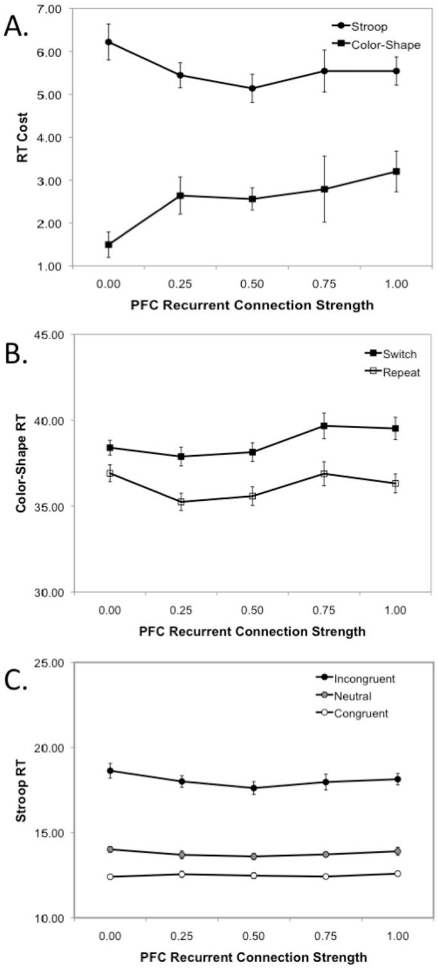

Figure 6.

Interference and Switch costs (A), and RTs in the (B) Color Shape task and (C) Stroop Task as a function of PFC recurrent connection strength. Bars indicate ± 2 standard errors, approximating 95% confidence intervals.

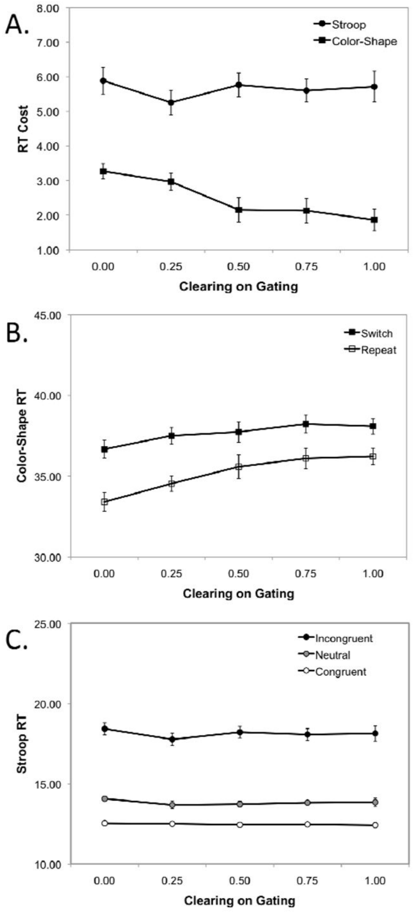

Figure 7.

Interference and Switch costs (A), and RTs in the (B) Color Shape task and (C) Stroop Task as a function of the clearing on gating parameter. Bars indicate ± 2 standard errors, approximating 95% confidence intervals.

Table 1

Correlations of Problems in a Longitudinal Twin Sample with Unity/Diversity EF Latent Variables at Age 17.

Correlated Variable	Common EF	Updating-Specific	Shifting-Specific
WAIS-IQ[a]	.51[*]	.49[*]	−.24[*]
Attention problems ages 7 to 14[b]			
Intercept of growth model	−.52[*]	.04	.20[~]
Slope of growth model[c]	.03	−.20	−.34[*]
Behavioral Disinhibition Age 12[d]			
Age 12 phenotypic	−.43[*]	.05	.14[~]
Age 12 genetic	−.58[*]	.07	.34[*]
Behavioral Disinhibition Age 17[d]			
Age 17 phenotypic	−.39[*]	.11	.11
Age 17 genetic	−.63[*]	.16	.30[*]
Self Restraint Age 14–36 months[e]			
Latent class membership phenotypic[f]	.29[*]	.02	−.21[*]
Latent class membership genetic	.49[*]	.01	−.34

[*] *p* < .05.

[~] *p* < .10.

[a] From Friedman et al. (2011).

[b] Reanalysis of Friedman et al. (2007).

[c] Higher values indicate greater decreases in problems across time.

[d] Reanalysis of Young et al. (2009).

[e] From Friedman et al. (2011).

[f] Cohen’s D to *r* conversions